Feature-Representation Transfer Learning for Human Activity Recognition

Ronald Mutegeki
School of Computer Science and Engineering
Kyungpook National University
Daegu, Korea
rmutegeki@knu.ac.kr

Dong Seog Han *
School of Electronics Engineering
Kyungpook National University
Daegu, Korea
dshan@knu.ac.kr*

***Abstract* — Many human-centered intelligent systems require information about the activities being performed by the user for the systems to function optimally. Human activity recognition (HAR) is at the core of such systems. Activity recognition requires vast amounts of labeled training data to perform adequately under a variety of circumstances. The lack of enough labeled training data led to transfer learning (TL), a phenomenon that uses knowledge learned from one task's dataset to easily perform a different task. In this paper, we show how TL can be used to improve the recognition of human activities with a small number of data samples. Using a convolutional neural network - long short-term memory (CNN - LSTM) deep learning ensemble classifier, we show how features learned from activities with motion enable us to easily learn features of stationery activities, even with a small dataset. TL improved model generalizability and reduced overfitting. To evaluate the performance, we used the UCI HAR dataset that contains 6 activities which was split into two sub tasks. The accuracy increased by over 4% whereas the loss decreased by 30% between a base model and the TL model. We also present the opportunities that using TL presents to the field of human activity recognition where new activities that have very small amounts of training data can be learned using data from already existing datasets.**

Keywords—Human activity recognition; transfer learning; ensemble model; feature-representation transfer learning

I. INTRODUCTION

In trying to identify the actions carried out by a person, studies in the human activity recognition (HAR) domain have been extensive. Due to the numerous areas of application which range from human-computer interaction technologies [1], healthcare, ambient assisted living, IoT, security among many others, research in this field has grown leaps and bounds in both academic and industry circles. The advent of deep learning technologies has intensified the search for a solution that can easily classify a user's activity in real-time thus being an active area of research. With these deep learning approaches requiring huge amounts of labeled training data to perform well, a phenomenon known as transfer learning (TL) [2] came to the rescue.

The TL phenomenon is not a brand-new way of acquiring knowledge. Cook *et al.* [3] illustrated how researchers in the artificial intelligence community have struggled for decades trying to build machines capable of matching or exceeding the mental capabilities of humans. These researchers are trying to design systems which can leverage experience and knowledge from previous tasks to easily learn a new task which has not been encountered before. Human learners appear to have inherent ways to transfer knowledge between tasks [4]. That is, we recognize and apply relevant knowledge from previous learning experiences when we encounter new tasks. The more related a new task is to our previous experience, the more easily we can master it. On the contrary, the majority of the common machine learning algorithms have traditionally addressed solving of tasks in isolation i.e. without referring to prior knowledge. With TL, we are spending less time to learn new tasks, less information is required of experts (usually human), and there is also an ability to handle more situations effectively. For HAR tasks, TL enables us to handle different sensor placement positions, different people performing similar activities differently, learning new activities from a dataset of different activities, among others. This ability to transfer knowledge learned from one dataset, known as the source, to another dataset, the target, is changing the way machine learning models deal with the scarcity of data.

For a machine learning model, training data is the currency on which performance of the model depends. With the recent advances in deep learning, researchers agree that the more data you have when training your model, the better your model will perform in testing and in real-life situations. Unfortunately, being able to collect and annotate the data needed to perform proper classification is a rather time consuming and costly task. Transfer learning is used in almost every deep learning model whenever the target dataset does not contain enough labeled data [5]. Whereas TL has been extensively researched for inductive learning, convolutional neural networks [5, 7] and reinforcement learning, this paper focuses on how TL can be applied to the human activity recognition domain with a more specific application to a convolutional neural network-long short-term memory (CNN-LSTM) ensemble classifier to observe the effects features/layers transferred from a source network have on the target network/model. In Section II, we explore previous works while Section III discusses the proposed method and experiment setup. Section IV discusses the obtained results and finally, Section V concludes this research work.

II. BACKGROUND AND PREVIOUS WORK

TL has been a subject of research for quite a while and advances have been made from basic transfer learning to deep learning based transfer learning techniques as surveyed by Cook *et al.* [3]. In this section, we give a generic understating of deep learning based transfer learning. We review some of the previous works in HAR, CNN-LSTM classifiers, and TL. Furthermore, we illustrate how these works differ from our proposed approach of using TL.

Human Activity Recognition Dataset

The current HAR problem aims at using sensors; accelerometer, gyroscope, magnetometer; that are built into IMU devices, and smartphones to recognize the activity being performed by the user of the device. The multimodal signals from these sensors are collected over time, which makes them time dependent. We therefore refer to the HAR problem as a time series classification (TSC) problem.

Ismail *et al.* [7] defined a time series $X = [x_1, x_2, \ldots, x_T]$, as an ordered set of real values and its length is the number of real values T. He goes further to define a dataset $D = \{(X_1, y_1), (X_2, y_2), \ldots, (X_N, y_N)\}$ as a collection of pairs (X_i, y_i) where X_i is a time series with y_i as its corresponding label or class. A model is then built to classify the dataset D so as to map the space of possible inputs X_i to a probability distribution over the class variable values Y_i which forms the foundation of any machine learning model for a TSC problem. Since HAR is a TSC problem, several deep learning and non-deep learning approaches have been suggested. To note a few; the multi-class SVM suggested by Anguita *et al.* [6], LSTM [10], CNN [7, 11], CNN-LSTM [12] among many others. The biggest challenge that many of these approaches face, especially the deep learning-based ones, has been with learning from very small datasets. This is typical of any deep learning model. The reason for small datasets stems from the difficulties associated with data collection and annotation. These small datasets lead the model to; overfitting on the training data, low accuracy and high losses.

The UCI human activity recognition version 1 dataset from [6] has been an outstanding benchmark for smartphone and wearable sensor-based activity recognition tasks. This is the dataset we use in this research to create the source and target datasets. A variety of work on HAR has been done based on this version of the open dataset [6, 10]. Whereas the goal of these works was to correctly classify the raw/processed signals into the 6 activities/classes, the goal of our research was to evaluate the effect of features learnt from 3 of the activities on learning the remaining 3 activities. We evaluate transfer learning for similarly distributed datasets but with different signal characteristics. More information about the dataset is provided in Section III.

Deep Learning for HAR

Machine learning has been at the core of research into human activity recognition for a very long time [6]. However, since AlexNet [9] won the ImageNet competition in 2012, deep learning has seen successful applications in a multitude of domains. Computer vision, natural language processing, speech recognition, etc. This success led several researchers to use various deep learning approaches in solving the HAR problem [8, 10, 11, 12].

Kim *et al.* [12] employed a CNN-LSTM hybrid model that consisted of 3 convolutional layers, 2 LSTM layers with 128 hidden units each and a softmax classifier for 2 classes. It is then validated on their own dataset [12] and the performance in terms of error and classification accuracy compared with other machine and deep learning models; SVM, LSTM, etc. In our experiments, we chose to use another implementation of the CNN-LSTM ensemble classifier that is explained in Section III, whose architecture and objectives clearly differ from [12].

TL for HAR

TL, in the context of deep learning, involves first training a base network on a source dataset and task, and then transfer the learned features (the network's weights) to a second network to be trained on a target dataset and task [7]. The absence of large amounts of labeled training data and difficulty to collect the same can pose a very serious challenge for any deep learning model. It's at this very juncture that transfer learning becomes very useful in the learning of features in similar domains or across different domains.

Bengio *et al.* [5] explored the layers at which a CNN is more general or specific for an image classification task. The extent to which a model generalizes on a dataset is the foundation on which transfer learning is built [5]. Throughout this paper, we will refer to the source dataset as the dataset we are transferring the pretrained model from, and the target dataset as the dataset we are transferring the pretrained model to. We adopted several TL lingua and techniques for model evaluation from [5]. However, our work significantly differs from [5]. In that the goal of our work is about transfer learning being applied to HAR, which is a time series problem and the model architectures we used are significantly different. We expound on this in Section III. The work by Ismail *et al.* [7] bears a distant similarity to our work in as far as being a TSC using transfer learning. Beyond that, it is different in such a way that whereas it evaluates the transferability of a fully CNN, our work explores TL for an ensemble CNN-LSTM classifier whose CNN layers are time distributed.

III. METHOD AND EXPERIMENT SETUP

In our experiments, we use the UCI human activity recognition dataset explained by Anguita *et al.* in [6]. This is a 6-activity dataset that contains the 3D (x, y, z) raw signals extracted from the accelerometer and gyroscope of a smartphone strapped to the waist of a subject [6]. Using the same dataset splitting technique described by Bengio *et al.*, the dataset was split into two subtasks; *task A* that contains data for activities involving movement i.e. *walking, walking_upstairs* and *walking_downstairs*. *Task B* contains stationery activities; *sitting, standing and lying*. The dataset is comprised of 7,352 train samples and 2,947 test samples. The size of data samples for tasks *A* (3,285, 1,387) and *B* (4,067, 1,560) (Train, Test) are evenly distributed in terms of the sample size and examples per activity.

We built a Keras sequential model that comprised of 2 1-D convolutional layers each with filters=64 and kernel_size =5, a 1-D maxpooling layer with pool_size=2 and a flatten layer which were all wrapped in a time distributed wrapper. This wrapper applies a layer to every temporal slice of an input to maintain the temporal nature of our data for the LSTM layer [13]. The feature map is used as input to an LSTM layer with 128 cells and a ReLu activation function. The output from the LSTM layer is then sent through a 100 unit fully connected (FC) dense layer with a ReLu activation before being classified by a 3-class softmax layer as shown in Fig. 1 below.

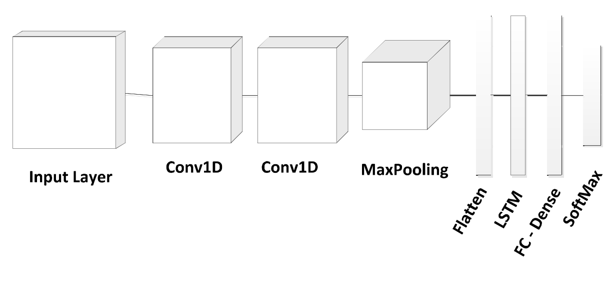

Fig. 1. A Keras sequential CNN-LSTM model to classify each of the tasks A and B into 3 classes.

To avoid overfitting and improve generalizability, due to the small size of the datasets being used, we applied a dropout layer between the FC and softmax layers. We used the adaptive moment estimation (ADAM) optimizer with its default settings and categorical_crossentropy as the loss function. The models were all trained for 15 epochs, minibatch size=64, and the dropout rate was set to 0.5.

IV. Results And Discussion

First the weights for models were initialized randomly. We pretrained a model using data from *task B* [5] and then transferred the layers of the model by creating a copy of the trained model as described by Bengio *et al.* The copied model was then re-trained with the same task's data to act as a control mechanism for the experiment just as described in [5] which we call the selffer model. The second model is the transfer learning model, where we pretrained a model initialized with random weights using data from *task A*. The model was saved and then retrained with *task B*'s data to evaluate how the model fared regarding transfer learning. As expected and just like the case in [5], we note that the performance of the model improved when the transferred feature layers are fine-tuned. This is different to when some or all layers are frozen. The models with transferred features performed and generalized a lot better than the models with randomly initialized weights. In addition to better performance, the loss and accuracy converged a lot faster for the models with transferred weights.

TABLE I below shows the average of the results attained from the various models we trained on the UCI human activity recognition dataset described in Section III. The base model was only trained on task B's data initialized with random weights without any layers being transferred attaining a 60% loss which is the worst of all the models. The selffer model achieves an almost similar performance with the base model since the data they train on is the same. The last model is the transfer learning model which achieves a far superior accuracy of almost 91%. The 28% error rate too was almost 2 times better than that of the other two models. This error rate shows that the TL model generalizes a lot better than the models that don't use TL. Regarding the layers that are transferable and those that are not, just like in [5], we observed that initializing the layer weights randomly performs dismally compared to using transferred weights.

TABLE I. Impact of Transfer Learning

	Accuracy	Loss
Base	86.42%	60.10%
Selffer	86.82%	56.85%
Transfer Learning	**90.80%**	**28.01%**

V. Conclusions

In this paper, we proposed a feature-representation transfer learning approach to human activity recognition using an ensemble CNN-LSTM model on the UCI HAR dataset. Whereas these ensemble CNN-LSTM models have been implemented for HAR problems, we demonstrated that TL can successfully be applied to these kinds of models. Fine-tuning the entire network on the target dataset would be the best approach after features have been transferred from the source dataset. Random weights perform poorly compared to transferred weights and improve generalizability of the model. TL works even for a small dataset and thus impacting performance of a model positively. For future work, TL will be applied to other public datasets with a richer set of activities and variations to the hyperparameters of the model. We shall also focus on the model's performance while applying TL

Acknowledgment

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT, R7124-16-0004, Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding).

References

[1] A. Iosifidis, A. Tefas, and I. Pitas, "Multi-view action recognition based on action volumes, fuzzy distances and cluster discriminant analysis," *Signal Processing*, vol. 93, no. 6, pp. 1445–1457, Jun. 2013.

[2] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," *ICML Workshop on Unsupervised and Transfer Learning*, pp. 17–36, 2012.

[3] D. Cook, K. D. Feuz, and N. Krishnan, "Transfer learning for activity recognition: A survey," *Knowledge Information Systems*, 36(3), pp. 537–556, 2013.

[4] L. Torrey and J. Shavlik, "Transfer Learning," In E. Soria, J. Martin, R. Magdalena, M. Martinez & A. Serrano, editor, *Handbook of Research on Machine Learning Applications*, IGI Global, 2009.

[5] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," Advances in Neural Information Processing Systems, 2014, pp. 3320–3328.

[6] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A Public Domain Dataset for Human Activity Recognition using Smartphones," *Proc. of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, Belgium, 24–26 April, 2013.

[7] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Transfer learning for time series classification," *IEEE Int. Conf. on Big Data*, pp 1367–1376, 2018.

[8] J. Wang, Y. Chen, H. Yu, M. Huang, and Q. Yang, "Easy transfer learning by exploiting intra-domain structures," *IEEE Int. Conf. on Multimedia and Expo (ICME)*, 2019.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, 25(2), pp. 1097–110, 2012.

[10] S. O. Eyobu and D. S. Han, "Feature Representation and Data Augmentation for Human Activity Classification Based on Wearable IMU Sensor Data Using a Deep LSTM Neural Network," *Sensors*, 2018, 18, 2892.

[11] F. M. Rueda, R. Grzeszick, G. A. Fink, S. Feldhorst and M. Hompel, "Convolutional Neural Networks for Human Activity Recognition Using Body-Worn Sensors," *Informatics,* 5(2), 26, May 2018.

[12] K. Kim, S. Choi, M. Chae, H. Park, J. Lee, and J. Park, "A Deep Learning Based Approach to Recognizing Accompanying Status of Smartphone Users Using Multimodal Data," *Journal of Intelligence and Information Systems*, vol. 25, no. 1, pp. 163–177, Mar. 2019.

[13] F. Chollet, "Layer wrappers," *Keras Documentation*, 2015, Online, Available: https://keras.io/layers/wrappers/#timedistributed, Accessed: Aug. 24, 2019.